Supplementary material:

SILT: Shadow-aware Iterative Label Tuning for Learning to Detect Shadows from Noisy Labels

Han Yang[1,*], Tianyu Wang[1,2*], Xiaowei Hu[3,1,†], and Chi-Wing Fu[1,2]

[1] The Chinese University of Hong Kong [2] The Shun Hing Institute of Advanced Engineering
[3] Shanghai Artificial Intelligence Laboratory

[*]Joint first authors
[†]Corresponding author (xwhu@cse.cuhk.edu.hk)

There are five parts in this supplementary material.

Part 1 presents additional demonstrations of the noisy labels in the SBU [4] and UCF [6] datasets.

Part 2 presents additional results of our Label Tuning method after 1 to 6 rounds of training.

Part 3 presents additional visual comparisons with the state-of-the-art methods.

Part 4 presents visual comparisons with two state-of-the-art methods trained on the original SBU [4] dataset and our tuned SBU dataset.

Part 5 presents additional statistics about SILT.

Part 1: Additional Demonstrations of Noisy Labels in the SBU and UCF Datasets

Figure 1. Images and corresponding noisy labels in the SBU [4] training set. We can see that some shadow details and some small shadows are missing from the original labels. Also, wrong labels may exist in both shadow and non-shadow regions.

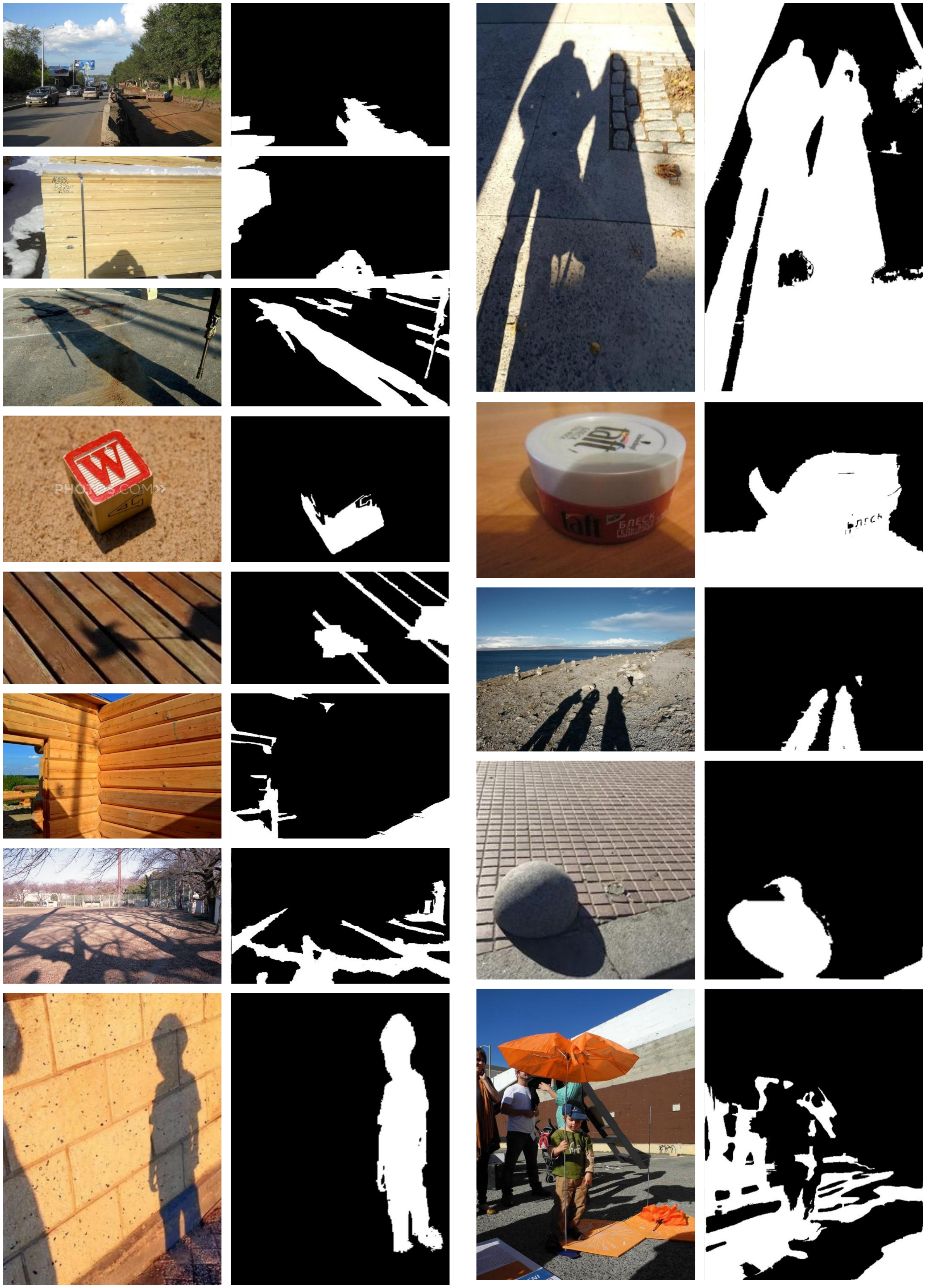


Figure 2. Images and corresponding noisy labels in the SBU [4] training set. We can see that some shadow details and some small shadows are missing from the original labels. Also, wrong labels may exist in both shadow and non-shadow regions.

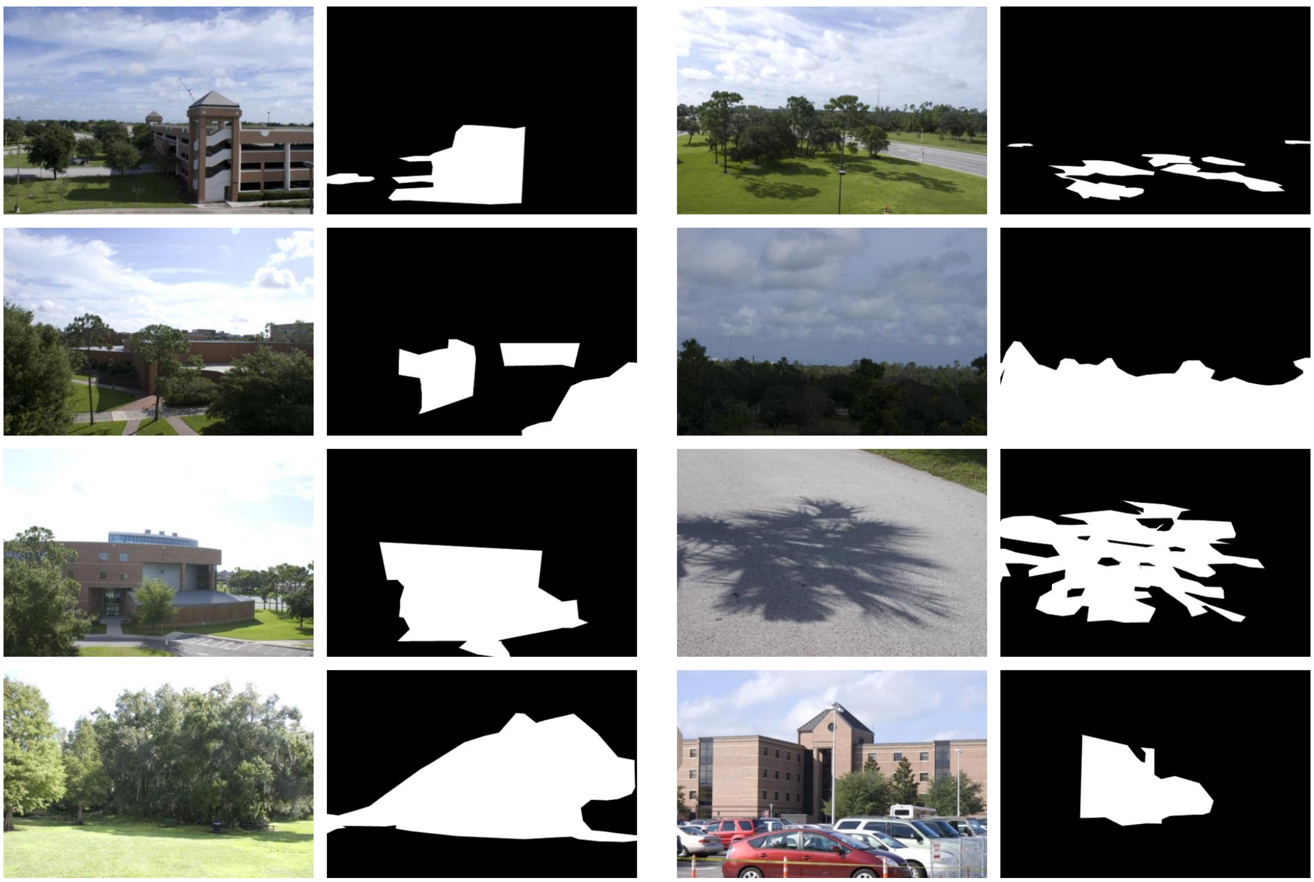


Figure 3. Images and corresponding noisy labels in the UCF [6] training set. We can see that the some shadow details and some small shadows are missing from the original labels.

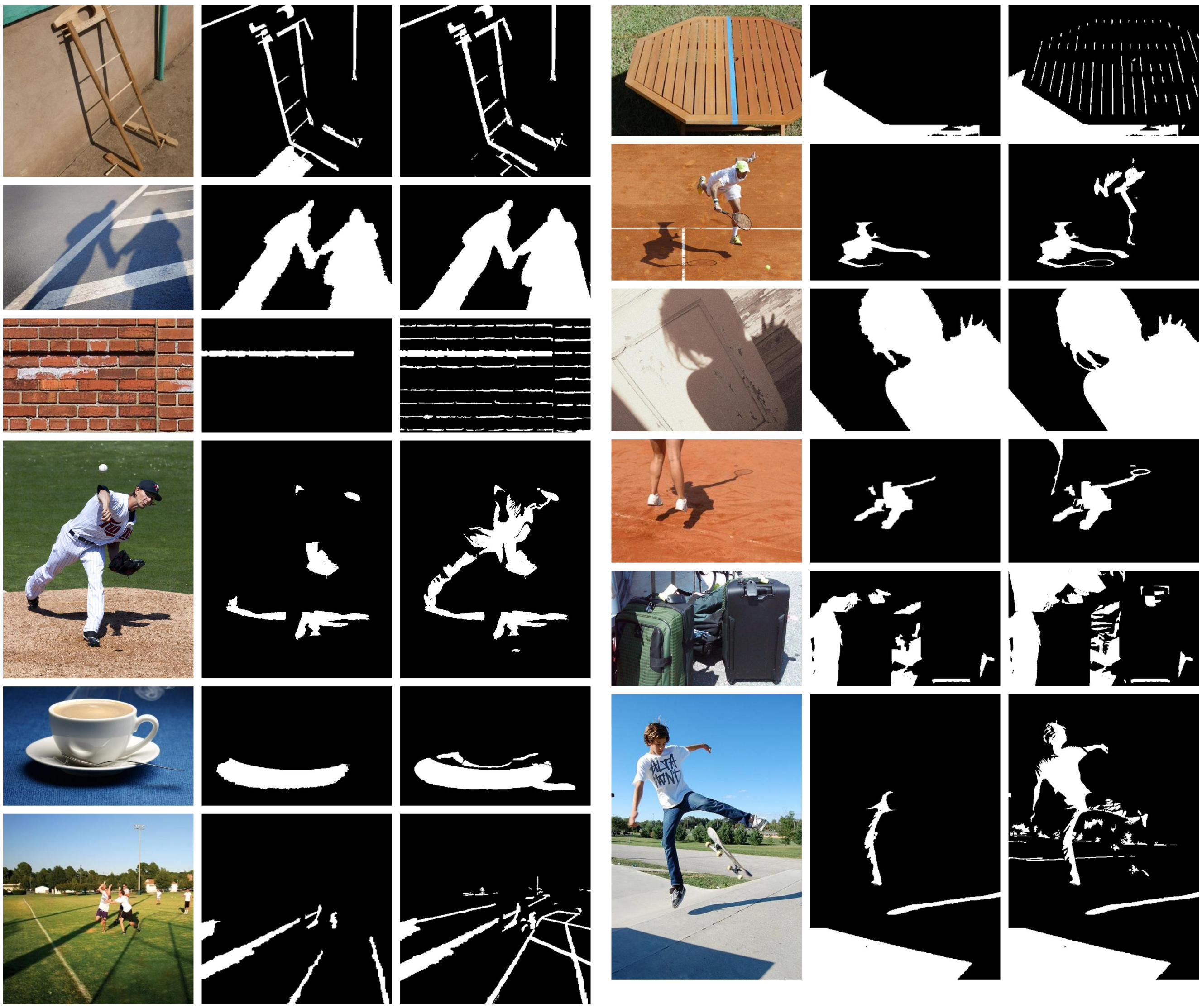
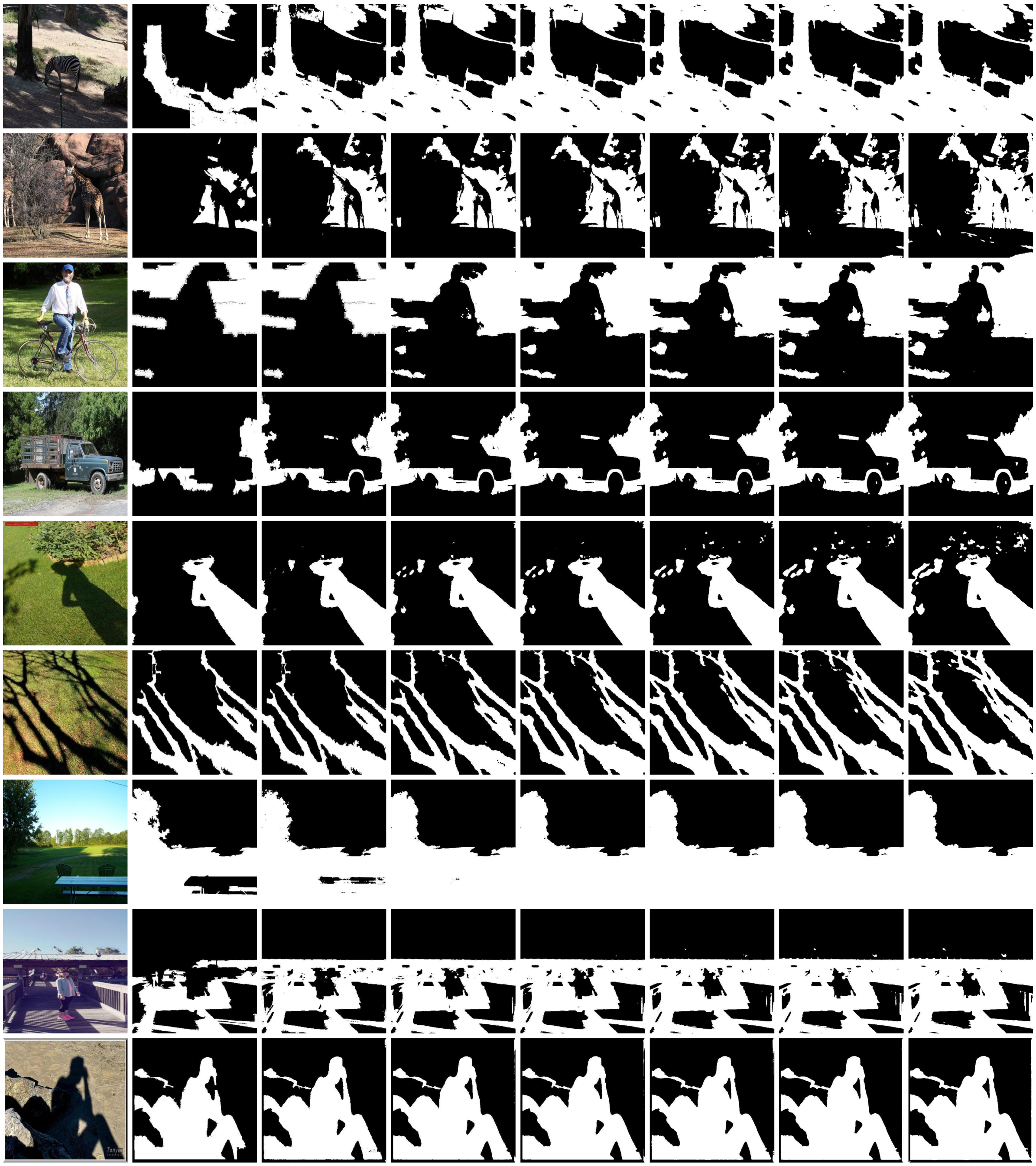


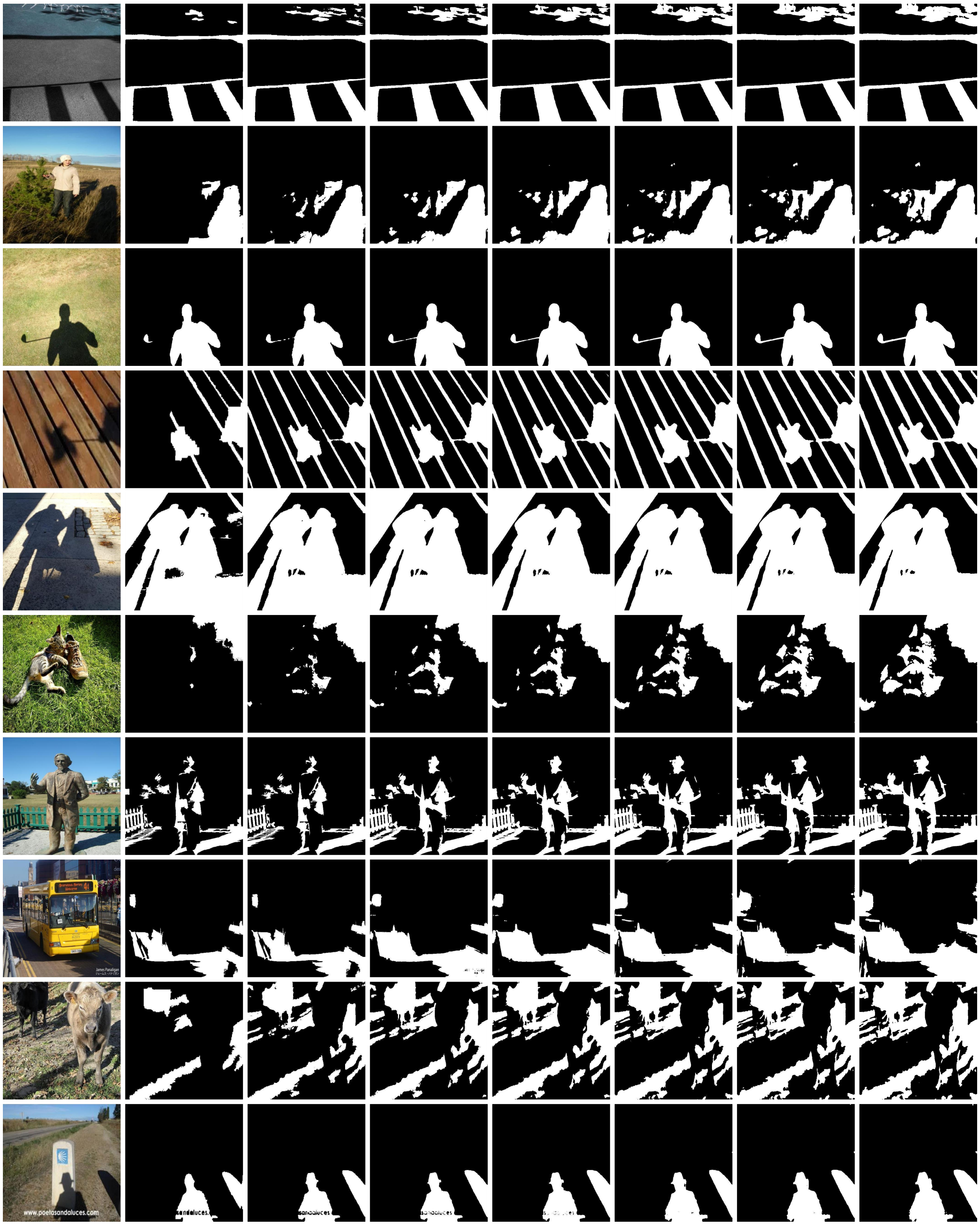
Figure 4. Images, original noisy labels, and our manually corrected clean labels in the SBU [4] test set.

Part 2: Additional Results Produced by Our Label Tuning



(a) Input images (b) Original labels (c) Round 1 (d) Round 2 (e) Round 3 (f) Round 4 (g) Round 5 (h) Round 6

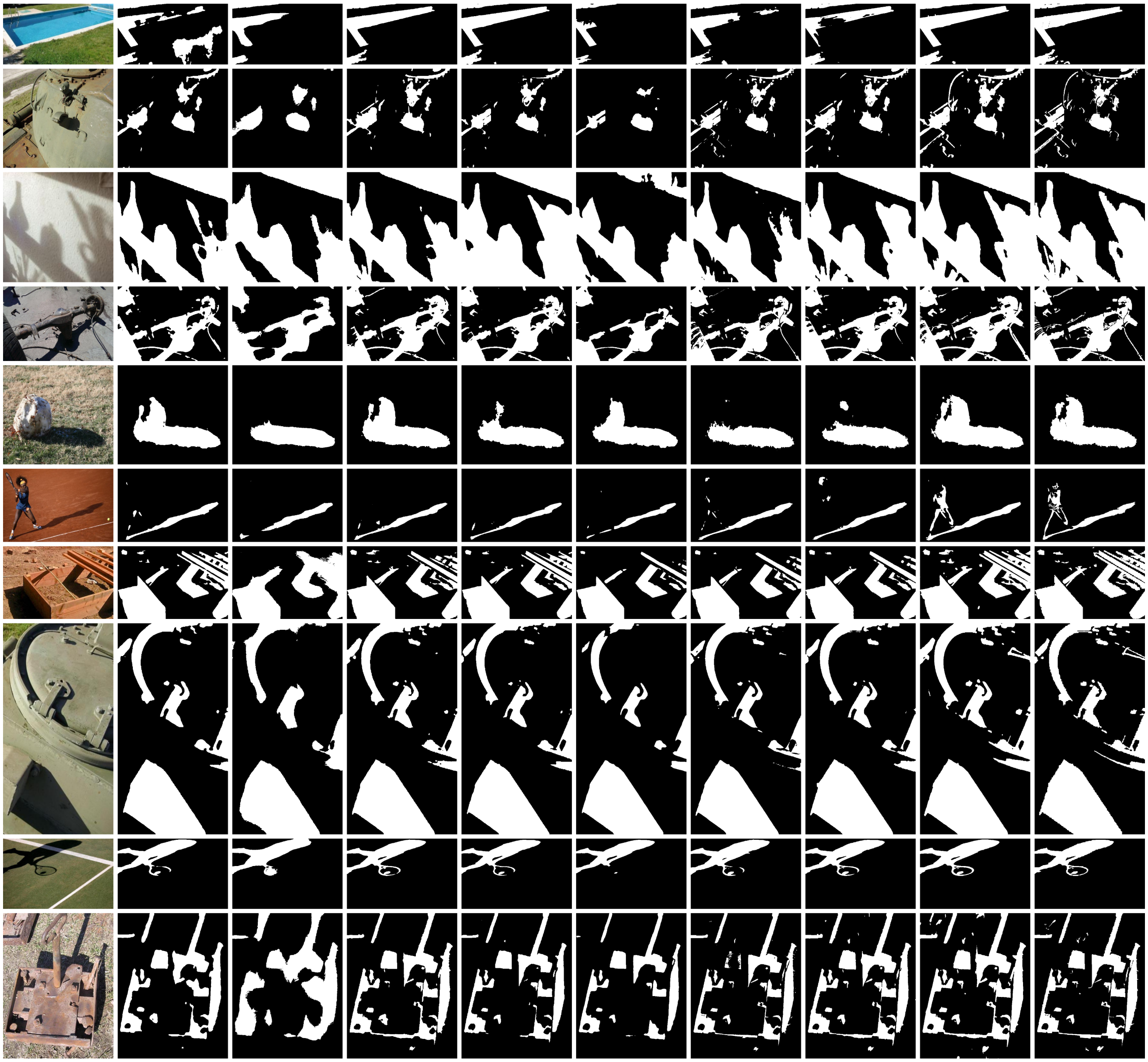
Figure 5. We show additional results of our Label Tuning after 1-6 rounds of training. We can observe that our Label Tuning can gradually bring back large missing shadows and add more details to the labels. Images are from the SBU [4] training set.



(a) Input images (b) Original labels (c) Round 1 (d) Round 2 (e) Round 3 (f) Round 4 (g) Round 5 (h) Round 6

Figure 6. We show additional results of our Label Tuning after 1-6 rounds of training. We can observe that our Label Tuning can gradually bring back large missing shadows and add more details to the labels. Images are from the SBU [4] training set.

Part 3: Additional Visual Comparisons with the State-of-the-Art Methods



(a) Input images | (b) BDRAR [7] | (c) DSC [3] | (d) DSD [5] | (e) MTMT [1] | (f) FSD [2] | (g) FDRnet [8] | (h) SDCM [?] | (i) SILT (ours) | (j) Ground truths

Figure 7. Visual comparison of shadow detection results produced by SOTA methods (b-h) and our SLIT (i). All SOTA methods are trained on the original SBU training set.

Part 4: Additional Visual Comparisons of the State-of-the-Art Methods Trained on Original SBU [4] Dataset and on Our Tuned Dataset

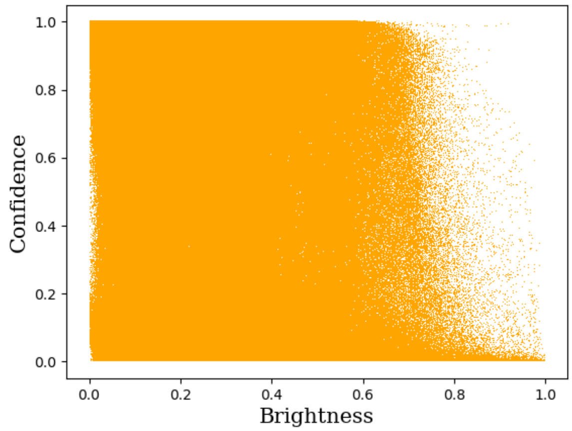
(a) Input images | (b) Using DSC [3] trained on the original SBU training set | (b) Using DSC [3] trained on our tuned SBU training set | (d) Ground truths

Figure 8. Visual comparison of shadow detection results produced by DSC [3] trained on the original SBU training set (b) and on our tuned training set (c). We can observe that DSC predicts more details and obtains better performance on self-shadow regions after trained on our tuned training set.

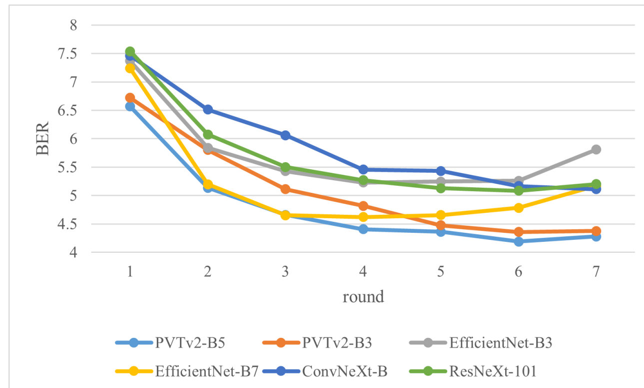
(a) Input images | (b) Using SDCM [?] trained on the original SBU training set | (b) Using SDCM [?] trained on our tuned SBU training set | (d) Ground truths

Figure 9. Visual comparison of shadow detection results produced by SDCM [?] trained on the original SBU training set (b) and on our tuned training set (c). We can observe that SDCM predicts more details and obtains better performance on self-shadow regions after trained on our tuned training set.

Part 5: Additional Statistics About SILT Framework



(a)



(b)

Figure 10. (a) This scatter plot displays the relationship between the brightness and confidence in the network's prediction of the shadow regions. Each dot in the plot represents a pixel in the predicted shadow masks. (b) This chart shows the BER after each round of self-training on SBU.

In Fig. 10 (a), we show the statistics of the brightness and confidence relationship in the network's prediction of the shadow regions. The plot reveals that the network has lower confidence in shadows with lighter colors, such as shadows cast on white walls. This observation leads us to set a lower threshold to encourage the network to label the shadows with light color.

In Fig. 10 (b), we show the BERs after each round of self-training on SBU [4] dataset. We can observe that, BERs first drop quickly due to the correction of large mislabeled regions, then drop further due to the refinement of fine details, and finally increase due to error accumulation.

References

[1] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. A multi-task mean teacher for semi-supervised shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5611–5620, 2020. 9

[2] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30:1925–1934, 2021. 9

[3] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7454–7462, 2018. 9, 10

[4] Tomás F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *European Conference on Computer Vision*, pages 816–832, 2016. 2, 3, 4, 6, 7, 8, 10, 12

[5] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5167–5176, 2019. 9

[6] Jiejie Zhu, Kegan G.G. Samuel, Syed Z. Masood, and Marshall F. Tappen. Learning to recognize shadows in monochromatic natural images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 223–230, 2010. 2, 5

[7] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *European Conference on Computer Vision*, pages 121–136, 2018. 9

[8] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *IEEE International Conference on Computer Vision*, pages 4702–4711, 2021. 9